

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第3919771号

(P3919771)

(45) 発行日 平成19年5月30日(2007.5.30)

(24) 登録日 平成19年2月23日(2007.2.23)

(51) Int. Cl.

G06F 17/28 (2006.01)

F I

G06F 17/28

Q

請求項の数 10 (全 23 頁)

<p>(21) 出願番号 特願2004-151966 (P2004-151966)</p> <p>(22) 出願日 平成16年5月21日(2004.5.21)</p> <p>(65) 公開番号 特開2005-108184 (P2005-108184A)</p> <p>(43) 公開日 平成17年4月21日(2005.4.21)</p> <p>審査請求日 平成16年7月7日(2004.7.7)</p> <p>(31) 優先権主張番号 特願2003-316236 (P2003-316236)</p> <p>(32) 優先日 平成15年9月9日(2003.9.9)</p> <p>(33) 優先権主張国 日本国(JP)</p> <p>(出願人による申告)平成16年度独立行政法人情報通信研究機構、研究テーマ「大規模コーパス音声対話翻訳技術の研究開発」に関する委託研究、産業活力再生特別措置法第30条の適用を受ける特許出願</p>	<p>(73) 特許権者 393031586 株式会社国際電気通信基礎技術研究所 京都府相楽郡精華町光台二丁目2番地2</p> <p>(74) 代理人 100099933 弁理士 清水 敏</p> <p>(72) 発明者 隅田 英一郎 京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内</p> <p>(72) 発明者 渡辺 太郎 京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内</p> <p>審査官 和田 財太</p>
--	--

最終頁に続く

(54) 【発明の名称】 機械翻訳システム、その制御装置、及びコンピュータプログラム

(57) 【特許請求の範囲】

【請求項1】

第1の言語の入力文に対しそれぞれ第2の言語の訳文を生成するための、複数の機械翻訳手段と、

前記複数の機械翻訳手段により生成される前記第2の言語の訳文をそれぞれ起点として訳文を変形し、当該変形後の訳文に対して訳文評価のための所定の評価基準に従って算出されるスコアが、変形前の訳文に対して前記所定の評価基準にしたがって算出されるスコアよりも改善されたものを出力するための訳文改良手段と、

前記訳文改良手段により出力された訳文のうちで前記評価基準により算出されたスコアが最良のものを前記入力文に対する訳文として選択するための訳文選択手段とを含む、機械翻訳システムであって、

前記訳文改良手段は、

入力される訳文に対し、所定のテキスト変形アルゴリズムによる所定の変形を行なって変形後の訳文を出力するための訳文変形手段と、

与えられる訳文に対して前記評価基準によるスコアの算出を行なうための訳文評価手段と、

前記訳文変形手段により変形された変形後の訳文に対して前記訳文評価手段により得られたスコアが、前記訳文変形手段による変形前の訳文に対して前記訳文評価手段により得られたスコアに対し改善されているか否かを判定し、改善が認められなくなるまで、前記訳文変形手段により変形された訳文に対する前記テキスト変形アルゴリズムによる変形、

10

20

及び当該変形前の訳文及び変形後の訳文に対する前記スコアの算出を繰返し実行するように前記訳文変形手段及び前記評価手段を制御するための繰返し制御手段とを含む、機械翻訳システム。

【請求項 2】

前記複数個の機械翻訳手段は、互いに異なる翻訳方式のものである、請求項 1 に記載の機械翻訳システム。

【請求項 3】

前記訳文変形手段は、一つの訳文に対して任意のテキスト変形アルゴリズムによる複数通りの変形を行なって複数個の変形後の訳文を生成するための手段を含み、

前記繰返し制御手段は、前記訳文変形手段により変形された複数通りの訳文の各々に対し、前記評価手段により算出されたスコアの改善が見られなくなるまで、前記訳文変形手段により変形された後の訳文に対し、前記変形、及び当該変形前の訳文及び変形後の訳文に対する前記スコアの算出を繰返し実行するように前記訳文変形手段及び前記評価手段を制御するための手段を含む、請求項 1 又は請求項 2 に記載の機械翻訳システム。

10

【請求項 4】

前記繰返し制御手段は、前記訳文変形手段により変形された複数通りの訳文のうち、前記評価手段により算出されたスコアが上位の所定個数内の各々に対し、前記評価手段により算出されるスコアの改善が見られなくなるまで、前記変形、及び当該変形前の訳文と変形後の訳文とに対する前記スコアの算出を繰返し実行するように前記訳文変形手段及び前記評価手段を制御するための手段を含む、請求項 1 又は請求項 2 に記載の機械翻訳システム

20

【請求項 5】

第 1 の言語の入力文に対しそれぞれ第 2 の言語の訳文を生成するための、複数個の機械翻訳手段と、

前記複数個の機械翻訳手段により生成される前記第 2 の言語の訳文をそれぞれ起点として訳文を変形し、当該変形後の訳文に対して訳文評価のための所定の評価基準に従って算出されるスコアが、変形前の訳文に対して前記所定の評価基準にしたがって算出されるスコアよりも改善されたものを出力するための訳文改良手段と、

前記訳文改良手段により出力された訳文のうちで前記評価基準により算出されたスコアが最良のものを前記入力文に対する訳文として選択するための訳文選択手段とを含む、機械翻訳システムであって、

30

前記訳文改良手段は、

入力される訳文に対し、所定のテキスト変形アルゴリズムによる変形を行なうための訳文変形手段と、

与えられた訳文に対して前記評価基準によるスコアの算出を行なうための訳文評価手段と、

前記訳文変形手段により変形された後の訳文に対する前記テキスト変形アルゴリズムによる変形、及び当該変形前の訳文と変形後の訳文とに対する前記スコアの算出を、所定の回数だけ繰返し実行するように前記訳文変形手段及び前記評価手段を制御するための繰返し制御手段とを含む、機械翻訳システム。

40

【請求項 6】

前記複数個の機械翻訳手段は、互いに異なる翻訳方式のものである、請求項 5 に記載の機械翻訳システム。

【請求項 7】

前記訳文選択手段は、前記繰返し制御手段による繰返しの中で得られた複数個の訳文のうち、前記訳文評価手段により算出されたスコアが最良のものを選択するための手段を含む、請求項 6 に記載の機械翻訳システム。

【請求項 8】

前記訳文評価手段は、与えられた訳文に対し、前記第 2 の言語の言語モデルと前記第 2 の言語から前記第 1 の言語への翻訳モデルとに基づいて尤度を算出し、当該与えられた訳文

50

に対するスコアとして出力するための手段を含む、請求項 1 ~ 請求項 7 のいずれかに記載の機械翻訳システム。

【請求項 9】

第 1 の言語の入力文に対しそれぞれ第 2 の言語の訳文を生成するための、複数の機械翻訳手段から前記第 2 の原語の訳文を受けるとして動作可能なコンピュータにより実行されると、当該コンピュータを、

前記複数の機械翻訳手段により生成される前記第 2 の言語の訳文をそれぞれ起点として訳文を変形し、当該変形後の訳文に対して訳文評価のための所定の評価基準に従って算出されるスコアが、変形前の訳文に対して前記所定の評価基準にしたがって算出されるスコアよりも改善されたものを出力するための訳文改良手段と、

10

前記訳文改良手段により出力された訳文のうちで前記評価基準により算出されたスコアが最良のものを前記入力文に対する訳文として選択するための訳文選択手段として機能させるコンピュータプログラムであって、

当該コンピュータプログラムは、

前記訳文改良手段を、

入力される訳文に対し、所定のテキスト変形アルゴリズムによる所定の変形を行なって変形後の訳文を出力するための訳文変形手段と、

与えられる訳文に対して前記評価基準によるスコアの算出を行なうための訳文評価手段と、

前記訳文変形手段により変形された変形後の訳文に対して前記訳文評価手段により得られたスコアが、前記訳文変形手段による変形前の訳文に対して前記訳文評価手段により得られたスコアに対し改善されているか否かを判定し、改善が認められなくなるまで、前記訳文変形手段により変形された訳文に対する前記テキスト変形アルゴリズムによる変形、及び当該変形前の訳文及び変形後の訳文に対する前記スコアの算出を繰返し実行するように前記訳文変形手段及び前記評価手段を制御するための繰返し制御手段として機能させる、コンピュータプログラム。

20

【請求項 10】

第 1 の言語の入力文に対しそれぞれ第 2 の言語の訳文を生成するための、複数の機械翻訳手段から、前記第 2 の言語の訳文を受けるとして動作可能なコンピュータにより実行されると、当該コンピュータを、

30

前記複数の機械翻訳手段により生成される前記第 2 の言語の訳文をそれぞれ起点として訳文を変形し、当該変形後の訳文に対して訳文評価のための所定の評価基準に従って算出されるスコアが、変形前の訳文に対して前記所定の評価基準にしたがって算出されるスコアよりも改善されたものを出力するための訳文改良手段と、

前記訳文改良手段により出力された訳文のうちで前記評価基準により算出されたスコアが最良のものを前記入力文に対する訳文として選択するための訳文選択手段として機能させるコンピュータプログラムであって、

前記コンピュータプログラムは、

前記訳文改良手段を、

入力される訳文に対し、所定のテキスト変形アルゴリズムによる変形を行なうための訳文変形手段と、

40

与えられた訳文に対して前記評価基準によるスコアの算出を行なうための訳文評価手段と、

前記訳文変形手段により変形された後の訳文に対する前記テキスト変形アルゴリズムによる変形、及び当該変形前の訳文と変形後の訳文とに対する前記スコアの算出を、所定の回数だけ繰返し実行するように前記訳文変形手段及び前記評価手段を制御するための繰返し制御手段として機能させる、コンピュータプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

50

この発明は機械翻訳システムに関し、特に、任意の2言語間の翻訳において、利用可能な言語リソースを生かしながら精度の高い翻訳を行なうことが可能な機械翻訳システムに関する。

【背景技術】

【0002】

社会経済活動のグローバル化の急速な進展に伴い、新たな言語対又は分野を対象とした機械翻訳システムを効率的に構築することが求められている。また、既に商用化され広く利用されている文字言語の翻訳においても、活発に研究され実用化が視野に入った音声言語の翻訳においても、現在の水準を越える高い翻訳品質が求められている。

【0003】

従来、機械翻訳システムの作成には、翻訳対象である2言語に堪能な専門家と年単位の時間と莫大な費用とを必要としている。そのような機械翻訳システムでは、現在求められている高い移植性又は品質を実現できない。これからは機械翻訳システムの作成にできるだけ人の関与をなくすようにし、機械化・工業化する必要がある。

【0004】

現在、世界の機械翻訳の研究においては、コーパスを利用する手法が従来手法の壁を破る成果を挙げつつある。コーパスを利用する機械翻訳の手法として、二つの代表的なアプローチがある。それらは、(1)用例翻訳と(2)統計翻訳とである。これらはいずれもコーパスを用いた半ば自動的な学習処理によって機械翻訳のためのシステムを構築できるという特徴をもつ。

【0005】

用例翻訳は、2言語の対訳を多数含む用例コーパスから翻訳パターンを抽出し、入力文に最も類似したパターンを利用して翻訳を行なう。第1の言語の入力文が与えられると、入力文に類似した第1の言語の文を対訳コーパスの中から探し出し、探し出された第1の言語の文の訳文(第2の言語)に基づいて出力文を作成する。

【0006】

一方、統計翻訳は、同様の用例コーパスから翻訳と言語との統計モデルを学習し、実行時にこれら二つの統計モデルに従って、確率を最大化するような訳文をサーチする。

【0007】

以下、従来技術の代表的な翻訳手法のうち、統計翻訳について説明し、さらに統計翻訳の精度をより高めようとする従来を試みを説明する。

【0008】

統計翻訳では、ある言語の文(今、この文を「J」と表す。)を他の言語の文(この文を「E」と表す。)に翻訳するという問題を、次の条件付確率を最大化する問題として定式化する。

【0009】

【数1】

$$\hat{E} = \arg \max_E P(E|J)$$

この式に対してベイズの定理を適用することにより次の式が得られる。

【0010】

【数2】

$$\hat{E} = \arg \max_E P(E)P(J|E)/P(J)$$

このうち、P(J)はEの算出とは無関係である(「E」は、上の式の左辺を表す)。従って次の式が得られる。

【0011】

【数3】

$$\hat{E} = \arg \max_E P(E)P(J|E)$$

10

20

30

40

50

右辺のうち第1項P(E)は言語モデルと呼ばれ、文Eの尤度を表す。第2項のP(J|E)は翻訳モデルと呼ばれ、文Eから文Jが生成される確率を表す。統計翻訳では、この確率が最大となるような訳文 \hat{E} をサーチしてそれを入力文に対する訳文とする。

【0012】

一方、こうした手法の限界を打破する方法として、チャンネルソース文の各単語に翻訳したものをチャンネルターゲット文の順番で並べたものを最初に生成し、この文に対し種々のオペレータを適用して多数の文を生成していく方法が提案されている(非特許文献1)。提案された手法では、その様に生成された文のうち、最も尤度の高いものを求めて訳文として選択する。

【0013】

【非特許文献1】ウルリヒ ゲルマン、ミハエル ヤール、ケビン ナイト、ダニエル マルキュ、ケンジ ヤマダ、「機械翻訳のための高速デコーディング及び最適デコーディング」、2001、ACL2001予稿集、ツールズ、フランス(Ulrich Ger mann, Michael Jahr, Kevin Knight, Daniel Marcu, and Kenji Yamada, "Fast decoding and optimal decoding for machine translation," (2001) in Proc. of ACL2001, Toulouse, France.)

【発明の開示】

【発明が解決しようとする課題】

【0014】

従来の用例翻訳と統計翻訳とのいずれの手法を用いたとしても、ある原理と言語データにおいて妥当な訳文を生成する枠組みを出るものではない。そのため、訳質をより高めようとするれば機械翻訳システムの内部そのものを変更しなければならず、期間、人手、費用のいずれの点でも改良に困難を伴う。

【0015】

また非特許文献1に記載の方法では、探索において局所的最適解に到達してしまうことが多いという問題があり、高精度の解が安定して得られる訳ではない。

【0016】

それに加えて、今後新たな翻訳手法が登場してきたとしても、各手法はその手法で完結したものとなり、それら新たな手法の限界を破った高い訳質の訳文を生成可能とするような枠組みは存在していない。

【0017】

それゆえに本発明の目的は、言語の組合せにかかわらず高品質の翻訳を得ることができる機械翻訳システムを提供することである。

【0018】

この発明の他の目的は、言語の組合せにかかわらず、ある程度の時間内に高品質の翻訳を得ることができる機械翻訳システムを提供することである。

【0019】

この発明のさらに他の目的は、利用可能な翻訳リソースを有効に活用可能で、言語の組合せにかかわらず、高品質の翻訳を安定して得ることができる機械翻訳システムを提供することである。

【課題を解決するための手段】

【0020】

本発明の第1の局面に係る機械翻訳システムは、第1の言語の入力文に対しそれぞれ第2の言語の訳文を生成するための、複数個の機械翻訳手段と、第2の言語の訳文をそれぞれ起点として、所定の評価方式に従った評価が改善するように訳文を改良するための訳文改良手段と、訳文改良手段により改良された訳文のうちで所定の条件を充足するものを入力文に対する訳文として選択するための訳文選択手段とを含む。

【0021】

10

20

30

40

50

複数の機械翻訳手段によって生成された訳文を改良し、改良された訳文の中で所定の条件を充足するものを入力文に対する訳文として選択する。単に機械翻訳手段の出力から訳文を選択するのではなく、それを改良するので、よりよい訳文が得られる。また複数の機械翻訳手段の出力結果を起点として訳文改良を行なうので、一つの機械翻訳手段の出力を使用する場合と比較して局所的最適解しか得られないというおそれが少なく、最適な訳文が得られる可能性が高い。また、既存の機械翻訳手段を使用することができる上、言語の組合せにも影響されずに訳質の高い訳文を得ることができる。

【 0 0 2 2 】

好ましくは、複数の機械翻訳手段は、互いに異なる方式のものである。

【 0 0 2 3 】

互いに異なる方式の機械翻訳手段を使用することで、最初に得られる訳文が互いに異なるものとなり、改良の過程で局所的最適解に陥る危険性が少なくなる。

【 0 0 2 4 】

より好ましくは、訳文改良手段は、入力される訳文に対し、所定の変形を行なうための訳文変形手段と、訳文変形手段により変形された訳文に対して評価を行なうための訳文評価手段と、訳文評価手段により変形された訳文に対して得られた評価が、変形前の訳文に対して得られた評価に対し改善されているか否かを判定し、改善が認められなくなるまで、訳文変形手段により変形された訳文に対し、変形及び評価を繰返し実行するように訳文変形手段及び評価手段を制御するための繰返し制御手段とを含む。

【 0 0 2 5 】

変形した訳文が、変形前の訳文に対して改善されているか否かを判定し、改善が認められなくなるまで変形と評価とを繰返す。複数の機械翻訳手段を使用することと併せ、局所的最適解に陥らずに大域的に見て最適な訳文が得られる可能性が高い。

【 0 0 2 6 】

さらに好ましくは、訳文変形手段は、一つの訳文に対して複数通りの変形を行なって複数の変形後の訳文を生成するための手段を含み、評価手段は、複数の変形後の訳文の各々に対し評価を行なうための手段を含む。

【 0 0 2 7 】

各訳文に対し複数通りの変形をして得られた訳文の各々が評価の対象となる。多くの訳文変形が評価及び変形の対象となるので、大域的に見て最適な訳文が得られる可能性が高い。

【 0 0 2 8 】

繰返し制御手段は、訳文変形手段により変形された複数通りの訳文の各々に対し、評価手段による評価の改善が見られなくなるまで変形及び評価を繰返し実行するように訳文変形手段及び評価手段を制御するための手段を含んでもよい。

【 0 0 2 9 】

評価の改善が見られなくなるまで繰返し変形及び評価が実行されるので、最終的には最も評価の高い訳文が得られる。

【 0 0 3 0 】

好ましくは、繰返し制御手段は、訳文変形手段により変形された複数通りの訳文のうち、評価が上位の所定個数内の各々に対し、評価手段による評価の改善が見られなくなるまで変形及び評価を繰返し実行するように訳文変形手段及び評価手段を制御するための手段を含む。

【 0 0 3 1 】

複数の訳文のうち、評価が上位の所定個数を選択し、変形及び評価を行なうことを評価の改善が見られなくなるまで繰返し実行する。評価が上位のもののみ選択するので、少ない計算量で大域的に見て最適な訳文が得られる可能性が高くなる。

【 0 0 3 2 】

より好ましくは、訳文改良手段は、入力される訳文に対し、所定の変形を行なうための訳文変形手段と、訳文変形手段により変形された訳文に対して評価を行なうための訳文評

10

20

30

40

50

価手段と、訳文変形手段により変形された訳文に対し、予め定められた回数だけ変形及び評価を繰返し実行するように訳文変形手段及び評価手段を制御するための繰返し制御手段とを含む。

【0033】

繰返しを、予め定められた回数だけ行なうので、所定時間内に確実に最終的な訳文が得られる。

【0034】

さらに好ましくは、訳文選択手段は、繰返し制御手段による繰返しの中で得られた複数個の訳文のうち、評価手段による評価が最も高かったものを選択するための手段を含む。

【0035】

最終的に評価が最も高い訳文を出力するので、評価基準に従った最も好ましい訳文を選択できる。

【0036】

訳文評価手段は、第2の言語の言語モデルと第2の言語から第1の言語への翻訳モデルとに基づいて訳文の尤度を算出するための手段を含んでもよい。

【0037】

最終的に言語モデルと翻訳モデルとに基づいて計算される尤度を評価値とし、尤度が最も高い訳文が得られる。そのため、得られる訳文は、自然な、訳質の高いものとなる。

【0038】

本発明の第2の局面に係るコンピュータプログラムは、コンピュータにより実行されると、当該コンピュータを上記したいずれかの機械翻訳システムとして動作させる。

【0039】

本発明の第3の局面に係る機械翻訳システムの制御装置は、複数個の機械翻訳装置に第1の言語の入力文を与え、それぞれ対応する第2の言語の訳文を得るための訳文取得手段と、訳文取得手段により得られた第2の言語の訳文を、各々が第2の言語の訳文をそれぞれ起点として、所定の評価方式に従った評価が改善するように訳文を変形するための複数個の訳文変形手段にそれぞれ与え、変形された複数個の訳文とそれぞれに付随する評価値とを受取るための変形訳文取得手段と、変形訳文取得手段が受取った訳文のうちで所定の条件を充足するものを入力文に対する訳文として選択し出力するための訳文選択手段とを含む。

【0040】

好ましくは、訳文選択手段は、変形訳文取得手段が受取った訳文のうちで最も高いスコアを持つものを選択するための手段を含む。

【0041】

本発明の第4の局面に係るコンピュータプログラムは、コンピュータにより実行されると、当該コンピュータを、機械翻訳システムの制御装置として動作させるためのコンピュータプログラムであって、機械翻訳システムは、第1の言語の文に対しそれぞれ第2の言語の訳文を生成するための、互いに異なる複数個の機械翻訳装置と、第2の言語の訳文をそれぞれ起点として、所定の評価方式に従った評価が改善するように訳文を変形するための複数個の訳文変形装置とを含み、制御装置は、第1の言語の入力文を複数個の機械翻訳装置に与え、対応する第2の言語の訳文を得るための訳文取得手段と、訳文取得手段により得られた第2の言語の訳文を、複数個の訳文変形装置にそれぞれ与え、変形された訳文とそれぞれに付随する評価値とを受取るための変形訳文取得手段と、変形訳文取得手段が受取った訳文のうちで所定の条件を充足するものを入力文に対する訳文として選択し出力するための訳文選択手段とを含む。

【発明を実施するための最良の形態】

【0042】

[第1の実施の形態]

本実施の形態の機械翻訳システムは、既存の翻訳リソースと、訳文の改良手法とを組合せた新たな枠組みに基づくものである。

10

20

30

40

50

【 0 0 4 3 】

- 構成 -

図 1 に、本実施の形態に係る機械翻訳システム 20 のブロック図を示す。図 1 を参照して、この機械翻訳システム 20 は、第 1 の言語（これを言語 J とする。）の入力文 30 を第 2 の言語（これを言語 E とする。）の訳文である出力文 42 に翻訳するシステムである。機械翻訳システム 20 は、第 1 の言語の入力文 30 を受け、後述するように種々の機械翻訳手法による訳文を訳文候補として生成するための訳文候補生成部 32 と、訳文候補生成部 32 により生成された訳文候補を、後述する方法に従って改良していき、所定の終了条件が充足された時点で最良の訳文候補を出力するための訳文改良部 36 と、訳文改良部 36 から改良された後の訳文候補が出力されたことに応答して、所定の終了条件が充足されているか否かを判定し、終了条件が充足されているときにはそのときまでに得られた改良後の訳文候補のうち、所定の評価基準に従って評価したスコアが最良のものを出力文 42 として選択し出力するための終了判定部 38 とを含む。

10

【 0 0 4 4 】

終了判定部 38 は、まだ終了条件が充足されていないと判定されると、訳文候補生成部 32 に対して再度初期候補を生成することを指示する制御信号 41 を送出する機能を持つ。訳文候補生成部 32 は、この制御信号 41 に応答して、先に初期候補を生成したときは異なる初期候補を生成し、訳文改良部 36 に与える機能を持つ。

【 0 0 4 5 】

図 2 に、訳文候補生成部 32 のより詳細な機能ブロック図を示す。図 2 を参照して、訳文候補生成部 32 は、与えられる文を翻訳し、それぞれ訳文 39A ~ 39E を出力するための第 1 ~ 第 5 の翻訳装置 35A ~ 35E と、入力文 30 を終了判定部 38 からの制御信号 41 に従ってこれら第 1 ~ 第 5 の翻訳装置 35A ~ 35E のうちのいずれか一つに分配するための分配部 33 と、第 1 ~ 第 5 の翻訳装置 35A ~ 35E のいずれかから出力される訳文を終了判定部 38 からの制御信号 41 に従って選択し、初期候補訳文 39 として出力するための選択部 37 とを含む。

20

【 0 0 4 6 】

第 1 ~ 第 5 の翻訳装置 35A ~ 35E は、本実施の形態では、互いに異なる方式により翻訳を行なう。従って、同一の入力文 30 が与えられたとしても、第 1 ~ 第 5 の翻訳装置 35A ~ 35E から得られる訳文 39A ~ 39E は互いに異なったものとなる可能性が高い。なお、この例では翻訳装置として 5 つを用いたものを示すが、この数は複数であればよく、5 に限定されるわけではない。また、同一の方式の翻訳装置であって、異なる翻訳知識を使用するものを用いても良い。

30

【 0 0 4 7 】

図 3 に、本実施の形態における第 1 の翻訳装置 35A の詳細なブロック図を示す。図 3 を参照して、第 1 の翻訳装置 35A は、第 1 の言語の文と第 2 の言語の訳文とからなる対訳文を多数含む対訳コーパス 34 と、対訳コーパス 34 を参照して、入力文 30 と対訳コーパス 34 の第 1 の言語の文の各々との類似の程度を示す類似尺度として tf/idf 基準 $P_{tf/idf}$ を算出するための tf/idf 算出部 50A とを含む。 tf/idf 基準 $P_{tf/idf}$ は、対訳コーパス 34 の第 1 の言語の各文を一つの文書とみなして、情報検索アルゴリズムで一般的に用いられる文書頻度と呼ばれる概念を用いて下の式により定義された尺度である。

40

【 0 0 4 8 】

【数 4】

$$P_{tf/idf}(J_k, J_0) = \sum_{i: J_{0,i} \in J_k} \frac{\log(N / df(J_{0,i})) / \log N}{|J_0|}$$

ただし J_0 は入力文、 $J_{0,i}$ は入力文 J_0 の i 番目の単語、 $df(J_{0,i})$ は入力文 J_0 の i 番目の単語 $J_{0,i}$ に対する文書頻度、 N は対訳コーパス 34 内の全ての訳文の数を、それぞれ示す。文書頻度 $df(J_{0,i})$ とは、入力文 J_0 中の i 番目の単語 $J_{0,i}$ が出現する文

50

書（本実施の形態では訳文）の数のことをいう。

【0049】

第1の翻訳装置35Aはさらに、対訳コーパス34に含まれている各対訳文（ J_k, E_k ）の第1の言語の文 J_k について、入力文 J_0 との間のDP（Dynamic Programming）マッチングを行なって編集距離 $dis(J_k, J_0)$ を算出するための編集距離算出部52Aと、 tf/idf 算出部50Aにより算出された tf/idf 基準 $P_{tf/idf}$ 及び編集距離算出部52Aにより算出された編集距離に基づいて、後述する式に従って各対訳文のスコアを算出するためのスコア算出部54Aを含む。

【0050】

編集距離算出部52Aにより算出される編集距離 $dis(J_k, J_0)$ は以下の式で表される。 10

【0051】

【数5】

$$dis(J_k, J_0) = I(J_k, J_0) + D(J_k, J_0) + S(J_k, J_0)$$

ただし k は $1 \leq k \leq N$ なる整数、 $I(J_k, J_0)$ 、 $D(J_k, J_0)$ 、及び $S(J_k, J_0)$ はそれぞれ文 J_0 から文 J_k に至るまでの挿入/削除/置換の数（編集距離）である。なお編集距離は、容易に入手可能なソフトウェアツールを利用して算出できる。

【0052】

スコア算出部54Aにより算出されるスコアは以下の式により表される。

【0053】

【数6】

$$score = \begin{cases} (1.0 - \alpha) \left(1.0 - \frac{dis(J_k, J_0)}{|J_0|}\right) + \alpha P_{tf/idf}(J_k, J_0) & (dis(J_k, J_0) > 0 \text{ のとき}) \\ 1.0 & (\text{その他のとき}) \end{cases}$$

ただし α は調整パラメータであり、本実施の形態では $\alpha = 0.2$ としている。

【0054】

図3を参照して、第1の翻訳装置35Aはさらに、スコア算出部54Aにより算出されたスコアに基づき、最も大きなスコアを持つ対訳文を選択し、その対訳文に含まれる第2の言語の文を第1の初期候補訳文39Aとして出力し図1に示す訳文改良部36に与えるための対訳文選択部56Aを含む。 30

【0055】

図4に第2の翻訳装置35Bの構成をブロック図形式で示す。図4を参照して、第2の翻訳装置35Bは、第1の言語である入力文30を、第3の言語の文に翻訳するための既存の翻訳システムからなる第1の中間翻訳装置50Bと、第1の中間翻訳装置50Bの出力する第3の言語の文を前述の第2の言語の文に翻訳するための第2の中間翻訳装置52Bとを含む。

【0056】

第1の中間翻訳装置50B及び第2の中間翻訳装置52Bとして性能のよいものが存在している場合、このように第3の言語を介在させて第1の言語から第2の言語に翻訳を行った結果、よい翻訳結果が得られる場合がある。本実施の形態のシステムでは、このような中間の言語を用いて得られた翻訳結果も、訳文の初期候補として利用できる。 40

【0057】

なお、第1及び第3の言語は互いに別の言語でもよいし、互いに同じ言語であってもよい。この場合、第1の中間翻訳装置50Bは第1の言語内での言換えを行なう装置ということになる。また、第2の言語と第3の言語とが同じ言語であってもよい。この場合、第2の中間翻訳装置52Bは、第2の言語内での言換えを行なう装置ということになる。

【0058】

図5に、第3の翻訳装置35Cの詳細なブロック図を示す。図5を参照して、第3の翻訳装置35Cは、入力文30をそれぞれ第2の言語に翻訳するための、互いに異なる翻訳 50

方式の第1～第3の翻訳部50C-1～50C-3と、これら第1～第3の翻訳部50C-1～50C-3の出力の訳質を所定の基準に従って評価し、その基準に従って最も優れていると思われるものを選択して第3の初期候補訳文39Cとして出力するための訳文選択部52Cとを含む。

【0059】

第1～第3の翻訳部50C-1～50C-3の翻訳方式は、互いに異なったものであればどのようなものであってもよい。

【0060】

訳文選択部52Cでの訳文の評価の基準としては、種々考えられるが、それらについては訳文改良部36での訳質評価の基準と共通していると考えられるので、ここではそれらの詳細な説明は行なわない。

10

【0061】

図6に、第4の翻訳装置35Dの詳細なブロック図を示す。図6を参照して、第4の翻訳装置35Dは、入力文30をそれぞれ第2の言語に翻訳するための、互いに異なる翻訳方式の第4～第6の翻訳部50D-1～50D-3と、これら第4～第6の翻訳部50D-1～50D-3の出力をマージして第4の初期候補訳文39Dとして出力するための訳文マージ部52Dとを含む。

【0062】

第4～第6の翻訳部50D-1～50D-3は、第1～第3の翻訳部50C-1～50C-3と同様、互いに異なる翻訳方式を用いるものであればどのようなものでもよい。

20

【0063】

訳文マージ部52Dによる訳文のマージとは以下のような処理である。簡単のために、入力文が「This is a pen.」という英語である場合を例として説明する。図7を参照して、この第4～第6の翻訳部50D-1～50D-3から、この入力文に対してそれぞれ「これはペンです」「これはペンだ」及び「これは筆です」という訳文が得られたものとする。訳文のマージでは、これら文を構成する単語又は単語列ごとに各訳文を比較し、最も多くの訳文に見出される単語又は単語列をマージ後の訳文の単語又は単語列として選択する。

【0064】

例えば、上記した図7に示す例では、枠60Dで示す部分は3つの訳文に共通している。従ってこの部分からは「これは」が訳文の要素として選択される。次に、枠61D及び62Dで示されるように、「ペン」という語が二つの訳文に見出され、「筆」という語は一つの訳文にしか見出されない。従ってこの部分からは「ペン」が訳文の要素として選択される。同様に枠63D～65Dからは、「です」が選択される。その結果、マージした後の訳文として枠69Dにより示す「これはペンです」が得られる。

30

【0065】

一般的にあって、複数の機械翻訳システムにおいて共通して使用されている単語又は単語列があれば、その単語又は単語列が訳語として妥当なものである可能性が高い。従って、上記したようなマージ処理によって、正しい訳文に近いものが得られる可能性が大きい。そのため、こうしたマージ処理の結果を初期候補訳文として利用する。

40

【0066】

図8に、第5の翻訳装置35Eの詳細なブロック図を示す。第5の翻訳装置35Eは、それぞれ入力文を第2の言語に翻訳するための第7～第9の翻訳部50E-1～50E-3と、第7～第9の翻訳部50E-1～50E-3の出力する訳文に対し、共有構造を持つ訳文を第5の初期候補訳文39Eとして生成するための訳文共有構造化部52Eとを含む。

【0067】

訳文の共有構造を生成する処理とは、次のような処理をいう。図9を参照して、図7に示すものと同様、入力文が「This is a pen.」という英語の文である場合を例として説明する。図9を参照して、この第7～第9の翻訳部50E-1～50E-3

50

から、この入力文に対してそれぞれ「これはペンです」「これはペンだ」及び「これは筆です」という訳文が得られたものとする。

【0068】

訳文の共有構造の生成は、基本的に訳文の単語列をグラフで表現する。例えば、図9の枠60Eで示すように互いに共通する部分（「これは」）はグラフでは一つのアーキにまとめる。そして、枠61Eと枠62E、及び枠63E～65Eによりそれぞれ表されるように、互に対応する部分であってかつ互いに異なる単語又は単語列が生成された部分では、それらの相違を別々のアーキ（「ペン」と「筆」、及び「です」と「だ」）で表現する。第5の初期候補訳文39Eはこのようなグラフ構造69Eを持った訳文候補である。

【0069】

本実施の形態では、以上の5つの翻訳装置を用いている。しかし、これ以外の翻訳システムであっても、第1の言語から第2の言語への翻訳を行なうことが可能なものであれば、どのようなものでも第1～第5の翻訳装置35A～35Eのいずれかに替えて、又はこれらに加えて、用いることができる。また第1～第5の翻訳装置35A～35Eも含め、利用可能な翻訳システムのどのような組合せでも訳文候補生成部32の要素として使用することが可能である。

【0070】

図10に、図1に示す訳文改良部36の詳細なブロック図を示す。図10を参照して、訳文改良部36は、訳文候補生成部32が出力した初期候補訳文39と、後述する訳文記憶部73から読出された訳文とのいずれか一方を選択するための訳文選択部70と、訳文選択部70の選択した訳文を、後述するように何らかの方式に従って変形するための訳文変形部71と、訳文変形部71により変形された後の訳文の訳質を所定の評価基準に従って評価しそのスコアを出力するための変形後訳文評価部72を含む。

【0071】

訳文改良部36はさらに、変形後訳文評価部72の出力するスコアとともに変形後訳文を記憶するための訳文記憶部73と、変形後訳文評価部72の出力するスコアに従って、訳文の改良を終了させるための終了条件が充足されているか否かを判定し、判定結果に従って繰返しを制御するための繰返し制御部74を含む。

【0072】

繰返し制御部74は、訳文記憶部73又は初期候補訳文39のいずれか一方を選択するように訳文選択部70に対する選択制御信号を送出する機能を持つ。ただし、処理の最初では訳文選択部70は必ず訳文39A～39Eを選択する。その後の処理で訳文39A～39Eを選択するか、訳文記憶部73の出力を選択するかは、どのようなスキームで訳文を変形させていくかにより異なってくる。

【0073】

繰返し制御部74はまた、変形後訳文評価部72によるスコアによっては終了条件が充足されていないと判定されたときには、訳文記憶部73に記憶された訳文のうちいずれか一つを所定の方式で選択して訳文選択部70に与えるよう訳文記憶部73を制御する機能、それと同時に訳文変形部71による訳文の変形を制御する機能、及び終了条件が充足されたと判定されたときには、後続の終了判定部38に対して訳文改良部36による訳文改良処理が完了したことを示す完了信号77を送出する機能も持つ。

【0074】

繰返し制御部74による訳文記憶部73からの訳文選択の順序は、訳文変形部71で行なう訳文変形の方式と関係して決定される。訳文変形部71で行なわれる訳文変形には、任意のテキスト変形アルゴリズムを利用することができる。本実施の形態では、訳文を統計翻訳で用いられる言語モデル及び翻訳モデルを用いてより尤度の高いものに変形する方式を使用する。

【0075】

テキスト変形アルゴリズムとしては、これ以外にも種々のものを用いることができる。以下にそれらを例示的に列挙する。

10

20

30

40

50

- (1) 言語モデルのみを用いて変形する。
- (2) 翻訳モデルのみを用いて変形する。
- (3) 予め人手で作成された文の書換えパターンに基づいて変形する。
- (4) 機械的に学習した書換えパターンに基づいて変形する。この際の学習としては、例えば機械翻訳の結果と用例コーパス中の正解とを比較し、その差分を変換パターンとして学習する方式が考えられる。
- (5) 単語の置換、挿入、削除等をランダムに、又は何らかのモデルに従って行なう。

【0076】

変形後訳文評価部72による訳質評価の手法としても、将来利用可能となるものも含めて種々の手法を利用することができる。本実施の形態では、統計的翻訳で用いられる言語モデル及び翻訳モデルによって訳文の尤度を計算し、変換後の訳文の尤度に改善が見られなくなったとき、又は所定の繰返し回数が完了したときに終了条件が充足されたと判定する。

10

【0077】

これ以外にも、訳質評価の尺度として考えられるものものを以下に例示的に列挙する。

- (1) 言語モデルのみにより得られる尤度
- (2) 翻訳モデルのみにより得られる尤度
- (3) 直訳度と呼ばれる尺度。直訳度としては、たとえば次の式で定義される Tanimoto 係数を使用できる。

【0078】

20

【数7】

$$Tanimoto \text{ 係数} = \frac{\text{原文に含まれる内容語の集合} \cap \text{翻訳文に含まれる内容語の集合}}{\text{原文に含まれる内容語の集合} \cup \text{翻訳文に含まれる内容語の集合}}$$

ただし |・| は集合の要素の数を表し、内容語とは、文の意味内容を判断する上で重要な語のことをいう。単語辞書にその単語が含まれているか否かによりその単語が内容語か否かを決定する手法が考えられる。

(4) 多重逆翻訳類似度。多重逆翻訳類似度とは、一旦得られた訳文を第1の言語への複数の翻訳システムで逆翻訳した結果が入力文とどの程度類似しているかを示す尺度である。この尺度が高ければ、訳文は入力文の正しい訳に近いと考えることができる。

30

(5) 参考訳を生成した上でこの参考訳を使用して訳文を評価する手法。この手法としては、BLEUSコア、WER (Word Error Rate)、NISTスコア、PER (Position Independent WER) 等、よく知られているものがある。以下に主なものについて説明する。

【0079】

<WER> Word-error-rate (単語誤り率) を表す。これは、基準となる模範的翻訳と比較した際の編集距離 (挿入 / 削除 / 置換) を反映したものである。

【0080】

<PER> 位置独立な (Position-independent) WER を表す。これは、位置的な問題を除き、挿入と削除とのみに関する編集距離を反映したものである。

40

【0081】

<BLEU> BLEUSコアを表す。これは翻訳結果のN-グラムのうち、基準となる模範訳の中に見出されるものの率を表す。上記したWER及びPERと異なり、BLEUSコアの値が高いほど翻訳の質は高い。

【0082】

なお、評価はこれ以外にもどのような手法を用いて行なってもよい。また、分野により特定の評価手法を採用するようにしてもよい。将来出現する有効な評価手法があれば、それを採用することも可能である。

【0083】

50

なお、繰返し制御部 74 は変形後の訳文の訳質に改善が見られない場合に繰返しを終了するが、訳質に改善が見られないときにもさらに変形を続けていくこともできる。ただし、本実施の形態では繰返し制御に山登り法を用いているので、訳質が劣化するようであれば繰返しは終了させる。

【0084】

この様にして訳文改良部 36 は訳文を変形していき、最も高い評価を持つ訳文を定めて出力文 76 としてスコアとともに終了判定部 38 に対して出力する。

【0085】

終了判定部 38 は、訳文改良部 36 からの出力文 76 とそのスコアとに基づいて、処理を終了させるか否かを判定する。本実施の形態では、単純に訳文候補生成部 32 に含まれる第 1 ~ 第 5 の翻訳装置 35A ~ 35E の出力の全てについて、訳文改良部 36 での処理が完了したか否かを判定する。全てについて完了したときにそれまでで最高のスコアを示した訳文を出力文 42 として出力する。全ての処理が完了していないときは、次の翻訳装置の訳文について上記した処理を実行するように制御信号を訳文候補生成部 32 に出力し、処理を続行させる。

【0086】

処理の終了条件としては、これに限らず以下に例示的に列挙する条件のうちの任意のものを採用することができる。ただし、この終了条件は訳質の改良の際の繰返し方法と関係があり、特定の繰返し方法には特定の終了判定しかできない場合、又は特定の繰返し方法には特定の終了判定が適用できないような場合等もあり得る。しかしそれらはいずれも設計事項に属する問題であり、当業者であれば適切な終了条件を選択できるであろう。

(1) 予め定められた繰返し回数又は計算時間を超えれば終了。

(2) 予め定められた繰返し回数又は計算時間内に訳質の改善が得られなくなれば終了。

(3) 訳質の改善が得られなくなれば終了。

(4) 予め定められた目標のスコアを得た訳文が得られれば終了。

【0087】

- 動作 -

機械翻訳システム 20 は以下の様に動作する。図 3 に示す対訳コーパス 34 には予め第 1 の言語の文と第 2 の言語の訳文とからなる多数の対訳文が含まれている。また言語モデル及び翻訳モデルも何らかの手段により予め準備されているものとする。

【0088】

図 1 を参照して、入力文 30 が訳文候補生成部 32 に与えられる。

【0089】

図 2 を参照して、分配部 33 が入力文 30 を第 1 の翻訳装置 35A に与える。

【0090】

図 3 を参照して、第 1 の翻訳装置 35A の tf/idf 算出部 50A は、入力文 30 と対訳コーパス 34 中の全ての対訳文のうち第 1 の言語の文との間で tf/idf 基準 $P_{tf/idf}$ を算出する。編集距離算出部 52A は、同様に入力文 30 と全ての対訳コーパス 34 中の第 1 の言語の文 J_k との間で編集距離 $dis(J_k, J_0)$ を算出する。

【0091】

スコア算出部 54A は、 tf/idf 算出部 50A により算出された tf/idf 基準 $P_{tf/idf}$ 及び編集距離算出部 52A により算出された編集距離 $dis(J_k, J_0)$ を用いて前述したスコア $score$ を次の式に従い算出する。

【0092】

【数 8】

$$score = \begin{cases} (1.0 - \alpha) \left(1.0 - \frac{dis(J_k, J_0)}{|J_0|} \right) + \alpha P_{tf/idf}(J_k, J_0) & (dis(J_k, J_0) > 0 \text{ のとき}) \\ 1.0 & (\text{その他のとき}) \end{cases}$$

対訳文選択部 56A は、対訳コーパス 34 に含まれる対訳文のうち、スコア $score$

10

20

30

40

50

の高いものを選択し、訳文 39 A として図 2 に示す選択部 37 に与える。

【0093】

選択部 37 は、終了判定部 38 からの制御信号に従って訳文 39 A を選択し、訳文 39 として訳文改良部 36 に与える。

【0094】

図 10 を参照して、訳文改良部 36 の訳文選択部 70 は、与えられた初期候補訳文 39 を選択して訳文変形部 71 に与える。訳文変形部 71 はこの訳文に対し所定のいく通りかの変形を行なって、変形後訳文評価部 72 に与える。変形後訳文評価部 72 はこの変形後の訳文の各々について前記したように所定の評価方式によって評価し、そのスコアとともに訳文記憶部 73 に与える。変形後訳文評価部 72 はまた、それらスコアを繰返し制御部 74 にも与える。

10

【0095】

繰返し制御部 74 は、これらのスコアが所定の条件を充足しているか否かを判定する。本実施の形態では、繰返し制御部 74 はいずれのスコアにも改善が認められなくなれば、又は所定回数だけ処理を繰返したときに、処理を終了する。最初の処理では通常はいずれかの変形により得られた訳文のスコアには改善が認められるので、繰返し制御部 74 は訳文選択部 70、訳文変形部 71、及び訳文記憶部 73 に対して処理の繰返しを指示するとともに、訳文記憶部 73 に対しては、先ほど記憶した訳文のうちでスコアの改善が見られた訳文のうちの一つを訳文選択部 70 に対し出力するように指示する。

【0096】

20

訳文選択部 70 は、今度は繰返し制御部 74 からの指示に従い、訳文記憶部 73 から与えられた変形後の訳文の一つを選択し、訳文変形部 71 に与える。訳文変形部 71 は、与えられた訳文に対し上記したのと同様のいく通りかの変形を行なう。この変形により得られた訳文の各々に対し変形後訳文評価部 72 が再度評価を行なってスコアを計算し、繰返し制御部 74 がスコアに改善が見られたか否かを判定する。訳文変形部 71、変形後訳文評価部 72、訳文記憶部 73、及び繰返し制御部 74 は、この処理を、いずれの訳文のスコアにも改善が見られなくなるまで、変形する対象の訳文を変えながら繰返し実行する。

【0097】

以上のように、一つの訳文候補に対し複数通りの変形を行なってそれらのスコアを評価し、スコアに改善が見られた訳文に対してさらに同様の変形及び評価を行なうことを、いずれの変形後訳文に対してもスコアの改善が見られなくなるまで（又は所定回数の繰返しが終了するまで）繰返し実行することにより、初期候補訳文 39 に比してスコアが大きく改善された訳文が得られる可能性が高くなる。

30

【0098】

どの訳文のスコアにも改善が見られなくなると（又は所定回数の繰返しが完了すると）、繰返し制御部 74 は以上の繰返し処理のうちで最もスコアが高かった訳文を出力文 76 として出力するように訳文記憶部 73 を制御し、あわせて完了信号を図 1 に示す終了判定部 38 に与える。

【0099】

終了判定部 38 は、この完了信号に応答して処理を終了させるべきか否かについて判定する。本実施の形態では、図 2 に示す第 1～第 5 の翻訳装置 35 A～35 E の全てについて、その生成した訳文に対する改良処理が終了して初めて全体の処理を終了する。従って終了判定部 38 は第 2 の翻訳装置 35 B により生成された訳文に基づいて、上記した訳文の改良処理を繰返すべく、訳文候補生成部 32 に対し制御信号 41 を与える。

40

【0100】

図 2 を参照して、分配部 33 はこの信号に応答して、入力文 30 を第 2 の翻訳装置 35 B に与える。第 2 の翻訳装置 35 B は、図 4 に示すように第 1 の中間翻訳装置 50 B と第 2 の中間翻訳装置 52 B とによる翻訳処理を行なって、訳文 39 B を生成し、選択部 37 に与える。

【0101】

50

選択部 37 は、終了判定部 38 からの制御信号に従って、第 2 の翻訳装置 35 B の出力する訳文 39 B を選択し、初期候補訳文 39 として訳文改良部 36 に与える。以下、訳文改良部 36 及び選択部 37 は第 1 の翻訳装置 35 A による訳文の処理と同様の処理を繰返す。

【 0 1 0 2 】

このようにして、第 1 ~ 第 5 の翻訳装置 35 A ~ 35 E が生成する訳文 39 A ~ 39 E の全てに対して前述した訳文改良処理が終了すると、図 10 に示す繰返し制御部 74 は、図 1 に示す終了判定部 38 に完了信号 77 を与える。終了判定部 38 はこの完了信号 77 を受けると、処理を終了すべき条件が充足されたと判定し、これまでの処理により得られた訳文のうち、最もスコアの高いものを出力文 42 として出力する。

10

【 0 1 0 3 】

訳文候補生成部 32 で用いられる翻訳装置としては、既存のもの、今後利用可能となるものも含め、どのようなものでも利用できる。

【 0 1 0 4 】

本実施の形態によれば、同一の入力文に対して互いに異なる複数の機械翻訳システムによって得られた訳文をそれぞれ始点として、訳文の改良を行ない、最良のスコアを持つ訳文をそれぞれ選択し、さらにその中で最も高いスコアのことを最終的な訳文として選択する。複数の訳文からスタートするので、局所解だけでなく、大局的な最適解が得られる可能性が高い。また、最初の訳文を得るための機械翻訳システムとしては、どのようなものでもよく、既存の機械翻訳システムを有効に利用することができる。さらに、今後開発される機械翻訳システム又は訳質評価の手法までも有効に利用することが可能で、この枠組みを使用してさらに訳文の品質を高くすることが可能である。

20

【 0 1 0 5 】

訳質評価のための基準とその手法、及び基本的な複数の機械翻訳システムさえ確立されていれば、言語の組合せを問わず、任意の言語の間での訳文の質を改善することができる。

【 0 1 0 6 】

さらに、上記した機械翻訳システムでは、訳文の品質の改善のために基本的には人手を必要とせず、システムの枠組みの開発も比較的簡単で短期で実現できるという特徴がある。

30

【 0 1 0 7 】

なお、上記した実施の形態では、変形後の訳文のうち、スコアに改善が見られたものの全てに対して再度、訳文の改良を行なう。しかし本発明はそのような実施の形態には限定されない。例えば、変形後の訳文でスコアに改善が見られたもののうち、上位の所定個数（例えば 1 個）の訳文に対してのみ、その後の変形と評価とを行なうようにしてもよい。

【 0 1 0 8 】

また、変形は複数通り行なうことが望ましいが、一通りの変形のみ使用するようによい。

【 0 1 0 9 】

[第 2 の実施の形態]

上記した様に、第 1 の実施の形態の装置はコンピュータにより実現することができる。また、図 2 等からも明らかなように、第 1 の実施の形態の装置は、その内部に互いに独立に動作可能な構成要素（例えば第 1 ~ 第 5 の翻訳装置 35 A ~ 35 E、第 1 ~ 第 3 の翻訳部 50 C - 1 ~ 50 C - 3、第 4 ~ 第 6 の翻訳部 50 D - 1 ~ 50 D - 3、第 7 ~ 第 9 の翻訳部 50 E - 1 ~ 50 E - 3 等）を含む。従って、コンピュータの通信機能とタスクの分配機能とを用いることにより、第 1 の実施の形態のシステムをネットワーク接続された複数のコンピュータにより実現することができる。第 2 の実施の形態のシステムは、そのように複数のコンピュータを互いにネットワーク接続して、上記した処理のうち並列に実行可能なものを個々のコンピュータで互いに並列に実行させるものである。

40

【 0 1 1 0 】

50

図 1 1 に、この機械翻訳システム 1 0 0 の概略の機能的構成を示す。図 1 1 を参照して、機械翻訳システム 1 0 0 は、それぞれ入力文 3 0 に基づいて、それぞれ別個の翻訳システムを用いて翻訳したものに基き、前記した訳文改良処理を行なって最良の訳文を生成するための複数個の最良訳生成部 1 0 2 A ~ 1 0 2 N と、これら最良訳生成部 1 0 2 A ~ 1 0 2 N により別個に生成された最良の翻訳の中でも最も高いスコアを有するものを選択して出力文 4 2 として出力するための訳文選択部 1 0 4 とを含む。

【 0 1 1 1 】

最良訳生成部 1 0 2 A ~ 1 0 2 N は、互いに別個のコンピュータと、その上で動作するプログラムにより実現できる。これらコンピュータとネットワーク接続されたホストコンピュータを設け、このホストコンピュータで各コンピュータに対する入力文 3 0 の分配、及び各コンピュータからの訳文の受信とその中での最良訳の選択とを行なうようにすればよい。

10

【 0 1 1 2 】

図 1 2 に、例として第 1 の最良訳生成部 1 0 2 A の機能的構成を示す。この最良訳生成部 1 0 2 A は、前述したようにホストコンピュータにネットワーク接続されたコンピュータと、その上で動作するプログラムとにより実現される。他の最良訳生成部も、初期候補の翻訳部が異なることを除き同様の構成を有する。

【 0 1 1 3 】

最良訳生成部 1 0 2 A は、図 2 に示す訳文候補生成部 3 2 と同様であるが、ただ一つの翻訳装置を持つ初期候補生成部 1 0 6 A と、初期候補生成部 1 0 6 A により生成された訳文を初期候補訳文として、図 1 0 に示す訳文改良部 3 6 と同様の処理を実行してこの最良訳生成部 1 0 2 A の出力文 1 0 8 A を生成しホストコンピュータに送信するための訳文改良部 1 0 7 A とを含む。

20

【 0 1 1 4 】

訳文改良部 1 0 7 A の機能的構成は、図 1 0 に示す訳文改良部 3 6 の構成と同様である。ただし、図 1 0 に示す訳文変形部 7 1 及び変形後訳文評価部 7 2 からなる処理は、同時並列的に行なうように構成することが可能である。そこで、これらの処理をネットワーク接続した他のコンピュータに同時並列的に行なわせる。

【 0 1 1 5 】

図 1 3 に、上記したコンピュータネットワークによる機械翻訳システムのネットワーク構成を概略的に示す。図 1 3 を参照して、この機械翻訳システムは、上記したシステムの動作全体を制御し、入力文を分配する処理及び訳文の中から最高スコアのものを選択する処理を行なうホストコンピュータ 2 0 0 と、入力文をホストコンピュータ 2 0 0 から受取り、互いに同時並列的に機械翻訳を行なって結果を初期候補訳文としてホストコンピュータ 2 0 0 に返信するための初期候補生成コンピュータ 2 1 0 A ~ 2 1 0 N と、ホストコンピュータ 2 0 0 からそれぞれ別の初期候補生成コンピュータにより生成された訳文を受取り、それを初期候補として訳文改良処理を実行するための訳文改良コンピュータ 2 2 0 A ~ 2 2 0 M とを含む。

30

【 0 1 1 6 】

このような構成の機械翻訳システムによれば、大量の計算を同時並列的に実行できるため、最終的な出力文が得られるまでの時間を大幅に短くすることができる。しかも得られる出力文の訳質及び応用の範囲は、第 1 の実施の形態のものと同様である。さらに訳文改良処理の中身を細かく分けることにより、さらに多くのコンピュータを用いて階層的に同時並列処理で実現することが可能になり、さらに処理の高速化を図ることが可能である。

40

【 0 1 1 7 】

[実施の形態の拡張]

上記した第 1 及び第 2 の実施の形態の構成に、さらに次に挙げるような機能を追加することができる。

【 0 1 1 8 】

(1) 上記した実施の形態の機械翻訳システムで得られた入力文 3 0 と出力文 4 2 との

50

対を記憶しておき、同じ入力文 30 に対しては同じ出力文 42 を返すようにする。この構成により、繰返し処理が不要になり、次回以降の処理において大幅な高速化が実現できる。

【0119】

(2) 上記した実施の形態の機械翻訳システムで得られた入力文 30 と出力文 42 との対を集め、対訳コーパスを拡張する。この拡張後の対訳コーパスを用いて、用例翻訳又は統計翻訳を再構築する。この拡張により、用例翻訳又は統計翻訳のカバーリッジ及び訳質を改善できる可能性が高い。

【0120】

- コンピュータによる実現 -

10

【0121】

なお、本実施の形態に係る機械翻訳システムは、コンピュータハードウェアと、当該コンピュータハードウェア上で動作するプログラム、及び当該コンピュータの記憶装置に記憶される対訳コーパス、翻訳モデル、及び言語モデルにより実現することもできる。

【0122】

そうしたプログラムは、上記した実施の形態の説明により、当業者であれば容易に実現できるであろう。

【0123】

図 14 は上記した機械翻訳システムを実現するコンピュータシステム 330 の外観を示し、図 15 はコンピュータシステム 330 の内部構成を示す。

20

【0124】

図 14 を参照して、このコンピュータシステム 330 は、FD (フレキシブルディスク) ドライブ 352 及び CD-ROM (コンパクトディスク読出専用メモリ) ドライブ 350 を有するコンピュータ 340 と、キーボード 346 と、マウス 348 と、モニタ 342 とを含む。

【0125】

図 15 を参照して、コンピュータ 340 は、FD ドライブ 352 及び CD-ROM ドライブ 350 に加えて、CPU (中央処理装置) 356 と、CPU 356、FD ドライブ 352 及び CD-ROM ドライブ 350 に接続されたバス 366 と、ブートアッププログラム等を記憶する読出専用メモリ (ROM) 358 と、バス 366 に接続され、プログラム命令、システムプログラム、及び作業データ等を記憶するランダムアクセスメモリ (RAM) 360 とを含む。コンピュータシステム 330 はさらに、プリンタ 344 を含んでいる。

30

【0126】

ここでは示さないが、コンピュータ 340 はさらにローカルエリアネットワーク (LAN) への接続を提供するネットワークアダプタボードを含んでもよい。

【0127】

コンピュータシステム 330 に上記した機械翻訳システムとしての動作を行なわせるためのコンピュータプログラムは、CD-ROM ドライブ 350 又は FD ドライブ 352 に挿入される CD-ROM 362 又は FD 364 に記憶され、さらにハードディスク 354 に転送される。又は、プログラムは図示しないネットワークを通じてコンピュータ 340 に送信されハードディスク 354 に記憶されてもよい。プログラムは実行の際に RAM 360 にロードされる。CD-ROM 362 から、FD 364 から、又はネットワークを介して、直接に RAM 360 にプログラムをロードしてもよい。

40

【0128】

このプログラムは、コンピュータ 340 に上記した各実施の形態の機械翻訳システムとしての動作を行なわせる複数の命令を含む。この方法を行なわせるのに必要な基本的機能のいくつかはコンピュータ 340 上で動作するオペレーティングシステム (OS) 又はサードパーティのプログラム、もしくはコンピュータ 340 にインストールされる各種ツールキットのモジュールにより提供される。従って、このプログラムはこの実施の形態のシ

50

システム及び方法を実現するのに必要な機能全てを必ずしも含まなくてよい。このプログラムは、命令のうち、所望の結果が得られるように制御されたやり方で適切な機能又は「ツール」を呼出すことにより、上記した機械翻訳システムを実現する命令のみを含んでいればよい。コンピュータシステム 330 の動作は周知であるので、ここでは繰返さない。

【0129】

今回開示された実施の形態は単に例示であって、本発明が上記した実施の形態のみに制限されるわけではない。本発明の範囲は、発明の詳細な説明の記載を参酌した上で、特許請求の範囲の各請求項によって示され、そこに記載された文言と均等の意味及び範囲内のすべての変更を含む。

【図面の簡単な説明】

10

【0130】

【図1】本発明の第1の実施の形態に係る機械翻訳システムの機能ブロック図である。

【図2】図1に示す訳文候補生成部32のより詳細な機能ブロック図である。

【図3】図2に示す第1の翻訳装置35Aの詳細な機能ブロック図である。

【図4】図2に示す第2の翻訳装置35Bの詳細な機能ブロック図である。

【図5】図2に示す第3の翻訳装置35Cの詳細な機能ブロック図である。

【図6】図2に示す第4の翻訳装置35Dの詳細な機能ブロック図である。

【図7】訳文のマージ処理を模式的に説明するための図である。

【図8】図2に示す第5の翻訳装置35Eの詳細な機能ブロック図である。

【図9】訳文の構造共有化処理を説明するための図である。

20

【図10】図1に示す訳文改良部36の機能ブロック図である。

【図11】本発明の第2の実施の形態に係る機械翻訳システムの機能ブロック図である。

【図12】図11に示す第1の最良訳生成部102Aの詳細な機能ブロック図である。

【図13】第2の実施の形態に係る機械翻訳システムのネットワーク構成を示す図である。

【図14】本発明の一実施の形態に係る機械翻訳システムを実現するコンピュータの外観を示す図である。

【図15】図14に示すコンピュータのブロック図である。

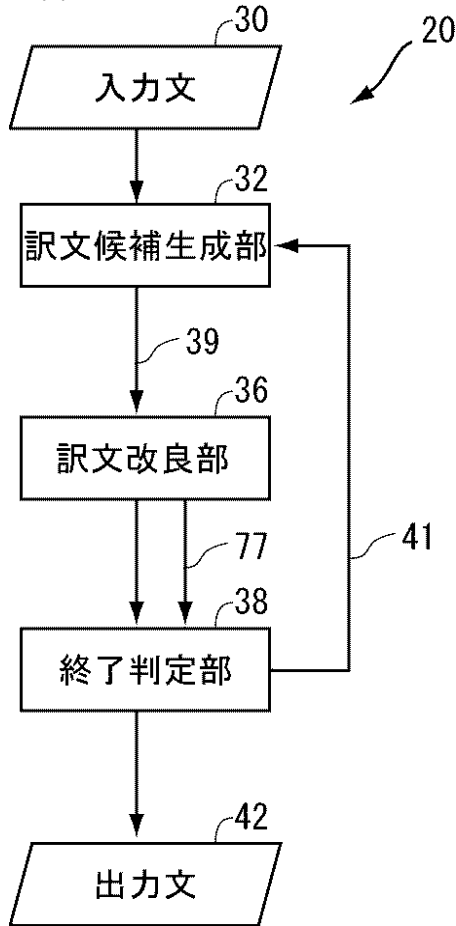
【符号の説明】

【0131】

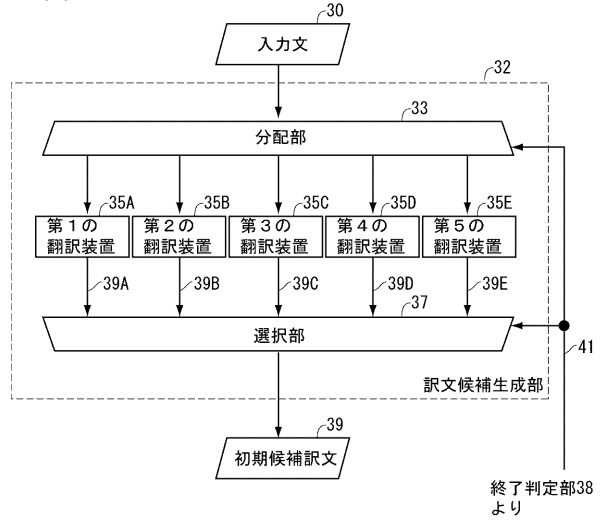
30

20, 100 機械翻訳システム、30 入力文、32 訳文候補生成部、33 分配部、34 対訳コーパス、36 訳文改良部、37 選択部、38 終了判定部、39 初期候補訳文、39A~39E 訳文、42 出力文、70 訳文選択部、71 訳文変形部、72 変形後訳文評価部、73 訳文記憶部、74 繰返し制御部、102A~102N 最良訳生成部、104 訳文選択部、106A 初期候補生成部、107A 訳文改良部

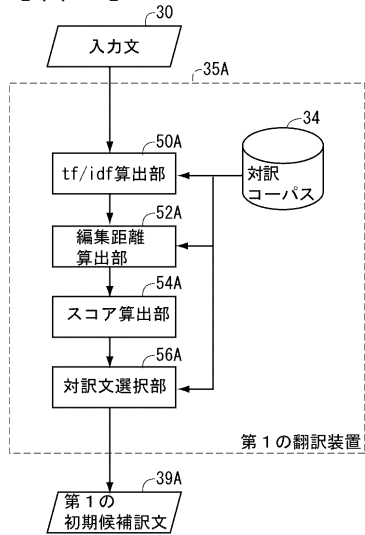
【図1】



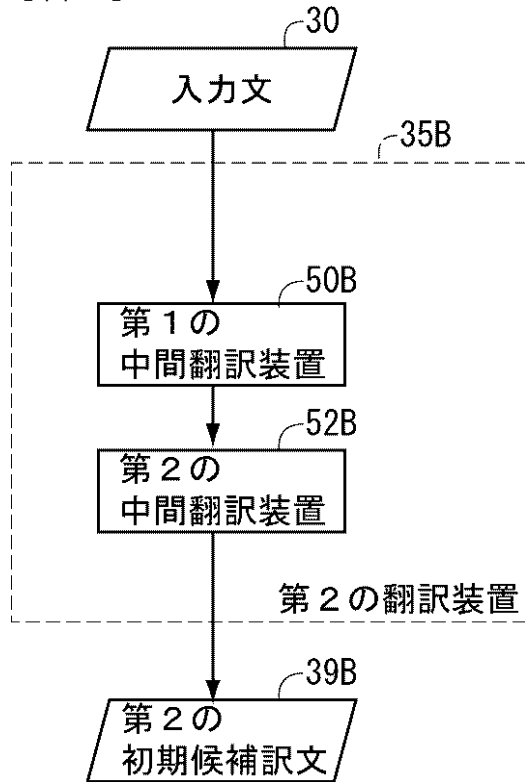
【図2】



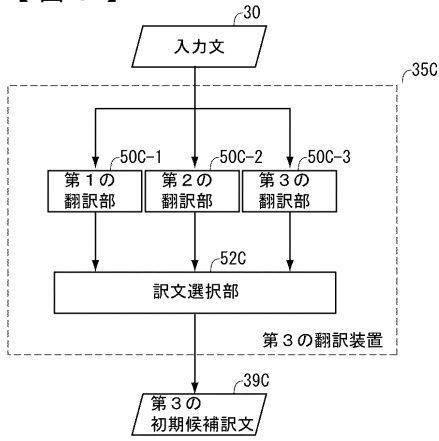
【図3】



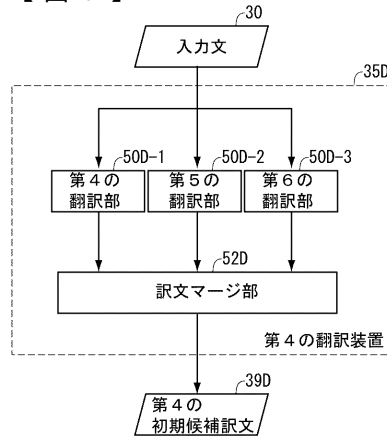
【図4】



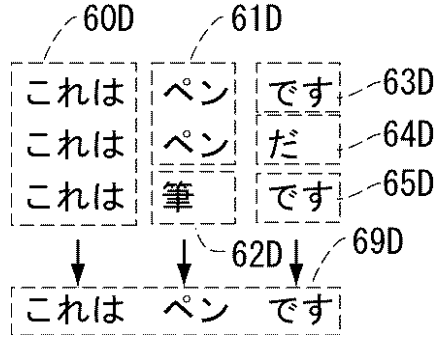
【図5】



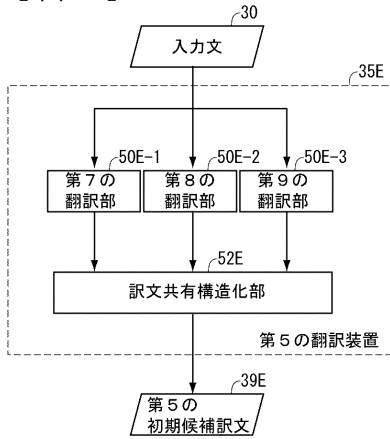
【図6】



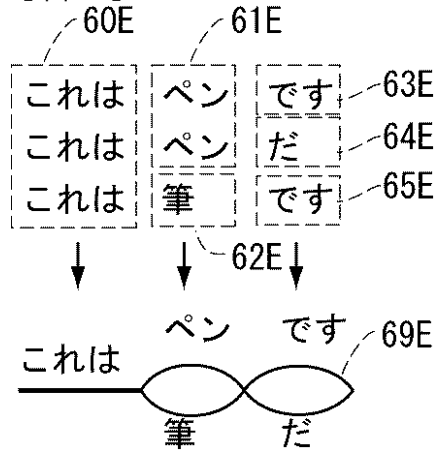
【図7】



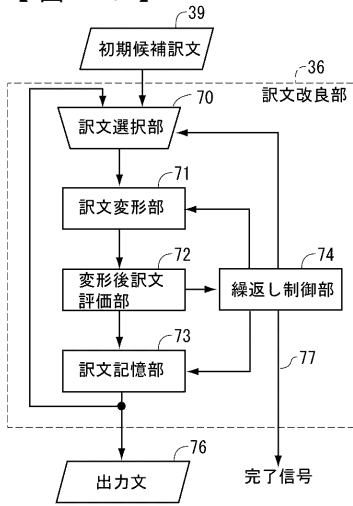
【図8】



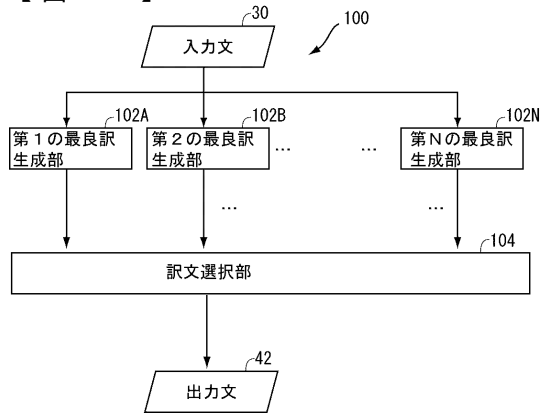
【図9】



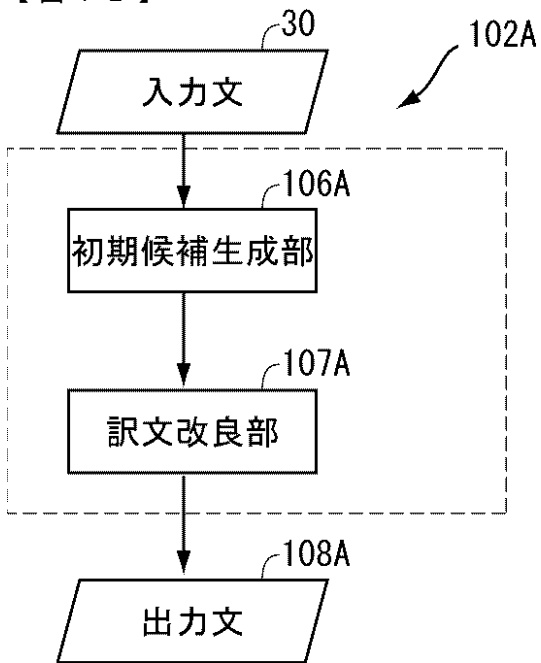
【図10】



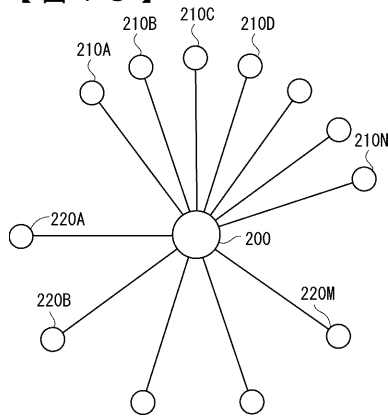
【図11】



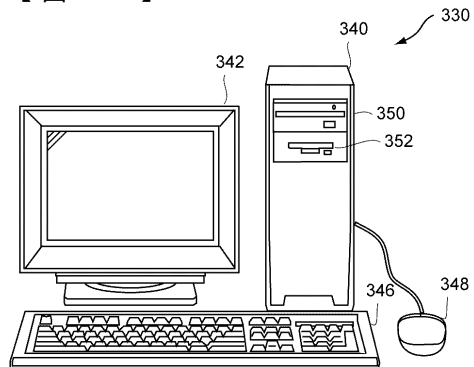
【図12】

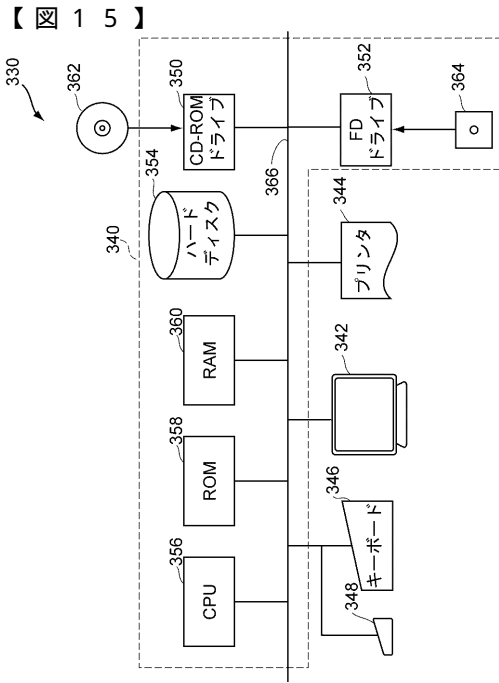


【図13】



【図14】





フロントページの続き

- (56)参考文献 速報型日英翻訳システムALTF LASH, 電子情報通信学会論文誌, 日本, 2001年 6月 1日, Vol. J84-D-II, No. 6, p. 1167 - p. 1174
速報型日英翻訳システムALFF LASH, NTT技術ジャーナル, 日本, 社団法人電気通信学会, 1999年 3月 1日, Vol. 11, No. 3, p. 81 - p. 83
SMT指標を用いた複数翻訳エンジンからの最適訳選択, 電子情報通信学会技術研究報告, 日本, 2002年 7月 8日, Vol. 102, No. 199, p. 63 - p. 68
SMT指標を用いた複数翻訳エンジンからの最適訳選択, 情報処理学会研究報告2002-NL-150, 日本, 2002年 7月16日, Vol. 2002, No. 66, p. 63 - p. 68

- (58)調査した分野(Int.Cl., DB名)

G06F 17/27 - 17/28